Generative models improve fairness of medical classifiers under distribution shifts

Year: 2023 | Citations: 133 | Authors: Ira Ktena, Olivia Wiles, Isabela Albuquerque, Sylvestre-Alvise Rebuffi, Ryutaro Tanno

---

Abstract

Domain generalization is a ubiquitous challenge for machine learning in healthcare. Model performance in real-world conditions might be lower than expected because of discrepancies between the data encountered during deployment and development. Underrepresentation of some groups or conditions during model development is a common cause of this phenomenon. This challenge is often not readily addressed by targeted data acquisition and 'labeling' by expert clinicians, which can be prohibitively expensive or practically impossible because of the rarity of conditions or the available clinical expertise. We hypothesize that advances in generative artificial intelligence can help mitigate this unmet need in a steerable fashion, enriching our training dataset with synthetic examples that address shortfalls of underrepresented conditions or subgroups. We show that diffusion models can automatically learn realistic augmentations from data in a label-efficient manner. We demonstrate that learned augmentations make models more robust and statistically fair in-distribution and out of distribution. To evaluate the generality of our approach, we studied three distinct medical imaging contexts of varying difficulty: (1) histopathology, (2) chest X-ray and (3) dermatology images. Complementing real samples with synthetic ones improved the robustness of models in all three medical tasks and increased fairness by improving the accuracy of clinical diagnosis within underrepresented groups, especially out of distribution. By generating synthetic image samples specific to underrepresented groups, diffusion models help medical image classifiers to achieve greater fairness metrics across a variety of medical disciplines and demographic attributes.